Emotion Recognition Using PHOG and LPQ features

Abhinav Dhall[1] Akshay Asthana[2] Roland Goecke[3,1] Tom Gedeon[1]
[1]School of Computer Science and [2]School of Engineering, CECS, Australian National University, Australia
[3]Vision & Sensing Group, Faculty of Information Sciences and Engineering, University of Canberra, Australia
abhinav.dhall@anu.edu.au, akshay.asthana@gmail.com, roland.goecke@ieee.org, tom.gedeon@anu.edu.au

Abstract— **We propose a method for automatic emotion recognition as part of the FERA 2011 competition. The system extracts pyramid of histogram of gradients (PHOG) and local phase quantisation (LPQ) features for encoding the shape and appearance information. For selecting the key frames, K-means clustering is applied to the normalised shape vectors derived from constraint local model (CLM) based face tracking on the image sequences. Shape vectors closest to the cluster centers are then used to extract the shape and appearance features. We demonstrate the results on the SSPNET GEMEP-FERA dataset. It comprises of both person specific and person independent partitions. For emotion classification we use support vector machine (SVM) and largest margin nearest neighbour (LMNN) and compare our results to the pre-computed FERA 2011 emotion challenge baseline.**

I. Introduction

We propose a method for automatic emotion recognition competition FERA 2011 [1] based on pyramid of histogram of gradients (PHOG) [2] and local phase quantisation (LPQ) [3] features. The experiment data set GEMEP-FERA [4] comprises of image sequences with actors uttering some monologues and exhibiting emotions, which in turn contains multiple units of individual facial expressions. Therefore, we quantise the number of frames in an image sequence using K-means clustering algorithm. Further we compare the emotion recognition results of out method with the baseline results provided with the GEMEP-FERA dataset[4]. For classification we experiment with support vector machines (SVM) [5] and largest margin nearest neighbour (LMNN) [6].

Automatic human emotion analysis finds a lot of application in HCI (Human Computer Interaction) Gaining insight into the state of the user's mind via facial emotion analysis can provide valuable information for affective sensing systems. Many psychological studies [7], [8] have discussed the importance of using multimodal system for accurate emotion analysis. Here we concentrate on automatic emotion analysis via facial expression recognition. Automatic emotion analysis finds applications in affective computing, intelligent environments, lie detection, psychiatry, emotion and paralinguistic communication and multimodal HCI.

Facial expression analysis has been an active field of research over the last two decades. It is mainly divided into two categories: image based and video based. In real world, human facial expressions are dynamic in nature. They constitute of an onset, one or more apex (peaks) and an offset. Studies [9], [10], have proven the effectiveness of video based facial expression analysis over the static analysis. In [9], Bassili suggests that motions cues from a face image sequence are enough to recognise an expression even with minimal spatial information.

In one of the early works Yacoob et al. [11], facial parts are tracked and optical flow is calculated at high gradient values of the image sequence. Here the head was static. The direction of the flow is quantised to eight levels in order to have a mid-level representation for high level facial expression classification. In [12], Black et al., use parametric models to extract parameters from facial features and use nearest neighbour classifier for FER. Different parametric models are used to differentiate between facial features relative to the head. In [13], a comparison is performed for various facial expression analysis techniques such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), optical flow and local filters such as Gabor wavelet representation via quantifying the Facial Action Coding System (FACS) [14]. In [15], Active Appearance Models are used for extracting facial features post fitting and machine learning techniques to classify emotions into FACS units [14]. Recently, [16] evaluate various state-of-the-art machine learning and image representation techniques on a new practical environment dataset for robust smile detection.

In their work, Pantic et al. [17] propose an automatic AU detection method for profile face image sequences. Face tracking is dealt with as a segmentation problem where the profile face is the foreground. It finds the largest connected component in HSV colour space and then use the watershed segmentation algorithm to finally segment the face. Using contour based method, 20 points are extracted which are then used to identify AU using a rule based method. Pantic et al.[18], propose two methods, first for automatic recognition of AU in video sequences and second for classifying AU coded expressions into learnt emotion categories. The method is suitable for analysing temporal sequence pattern. Post face registeration, a temporal template called Motion History Image (MHI) [19] are constructed from the image sequence. Then temporal rules are used to identify AUs in the Cohn-Kanade [20] and the MMI [21] databases. The method achieves approximately 90% recognition rate when detecting 27 AUs. Further the work is extended in [22], where a wavelet based gentleboost template is used to track 20 facial fiducial points, which further are used to construct spatio-temporal features. A subset of features is selected using AdaBoost and SVM is used for checking the presence of

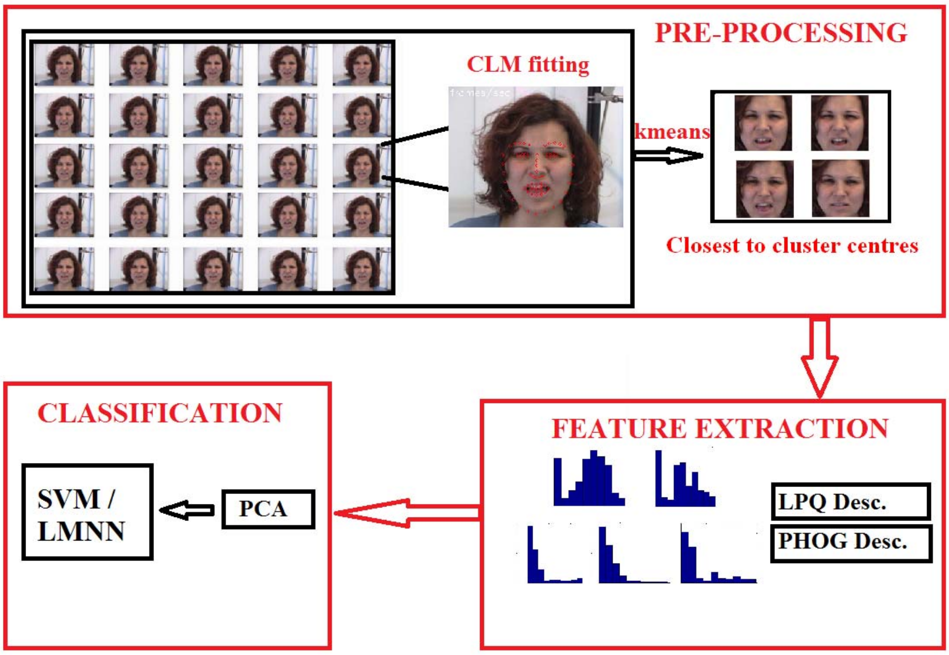

Fig. 1. Schematic representation of the process of testing sequences in our method.

AUs.

In [23], the authors propose volume LBP (VLBP) and LBP-TOP for dynamic facial expression analysis using LBP descriptors. In [24], the authors propose the use of LPQ for static facial expression analysis and extend LPQ to LPQ-TOP with the concept similar to LBP-TOP for extracting temporal information. In [25], a system based on PHOG and bag of words BOW [26] features for robust static facial expression analysis is used. We use PHOG features in our temporal emotion analysis system. The novelty of the GEMEP-FERA dataset is that there are no pre-defined onset, apex and offsets and the actors are speaking sentences, in which they exhibit same expression multiple times. The dataset has varied pose, dynamic head movements and occlusion. Our method gives good result on this dataset.

The rest of the paper is organised as follows: Section II presents the proposed technique in detail. Section III discusses the experimental results for emotion recognition of our method on GEMET-FERA dataset. Finally, Section IV provides the conclusions and future work.

II. System

Our emotion recognition system starts with face tracking using constraint local models (CLM) [27]. The shape vectors of the face in an image sequence are then normalised. The K-means clustering algorithm is applied to the normalised shape vectors and images having face shape vectors closest to the cluster centres are chosen for further processing. Further, the Viola-Jones [28] face detector is applied to the chosen images. Then we compute the PHOG and LPQ features on the cropped faces. For classification we use SVM and LMNN. Figure 1 descibes the test process of our method. The system is explained is described in depth in the following sub sections.

1) Face tracking using CLM: In recent times, learnt model-based techniques have been extensively used in non-rigid deformable object fitting. We use the CLM [27] for face tracking in the image sequences. It is based on fitting a parameterised shape model to the location landmark points of the face. It predicts locations of the model's landmarks by utilising an ensemble of local feature detectors, which are then combined by enforcing a prior over their joint motion. The distribution of the landmark locations is represented non-parametrically and optimised via subspace constrained meanshifts. It fits well to various poses. We used a person independent model which was trained on the Multi-PIE database [27]. Using the shape parameters from the fitted model a rough region of interest around the face is extracted. Figure 2 shows a snapshot of the face tracking using CLM. The fitting process gives a row vector P containing location of each of the n landmark points

$$P = [x_1; y_1 \dots x_n; y_n] \tag{1}$$

P is then normalised by taking the horizontal Euclidean distance between the outer eye corners on the left and right side. The vertical normalisation distance is the Euclidean distance between the tip of the nose and the midpoint between the eyebrows. We denote the normalised shape vector as P^n.

2) Clustering based sequence quantisation: P^n are calculated for all the images in an sequence. As the amount of motion in two consecutive frames is very sparse, we wish to remove the redundant frames so that the features are extracted on the frames which efficiently describe the

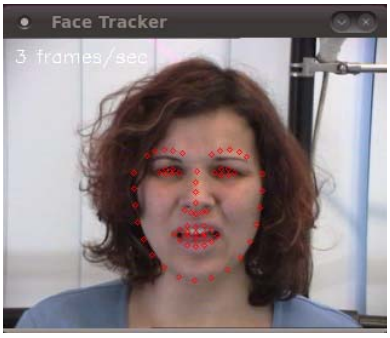

Fig. 2. The figure is a snapshot for CLM based face tracking. (taken from FEEDTUM [30] database).

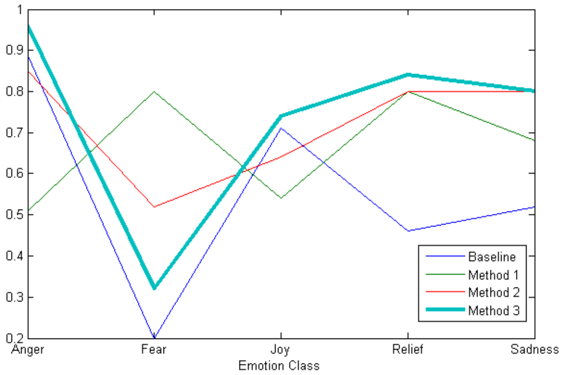

Fig. 3. The graph depicts the comparison of the four methods.

temporal dynamics of the expression. Therefore for selecting key frames, we compute K-means clustering algorithm on the normalised shape vectors. In [29], the authors also perform K-means clustering for computing their similarity features. However it is different from our approach as they compute it on the apex images, in order to divide them into various clusters. Our aim is different here, we want to remove the redundant frames. In addition the image sequences in the GEMET-FERA database, have multiple apex in an image sequence and exhibit an expressions multiple times. Post calculating the cluster centers we search for the nearest neighbour of the cluster centers. The algorithm is summarised in Algorithm 1.

Post searching for the nearest neighbour of the cluster centers the Viola-Jones face detector [28] is applied to the images corresponding to those nearest neighbours. Further we extract the features on the cropped faces computed by the face detector.

A. Shape feature extraction using PHOG

For extracting shape information we use PHOG [2] features. PHOG is a spatial pyramid extension of the histogram of gradients (HOG) [31] descriptors. The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image and has been used extensively in computer vision methods. PHOG features being an extension of HOG have shown good performance in object recognition

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	10	3	7	0	2
Fear	1	10	0	2	0
Joy	2	2	12	0	1
Relief	1	0	1	11	2
Sadness	0	0	0	3	10

TABLE I

CONFUSION MATRIX FOR PERSON INDEPENDENT TEST FOR OUR METHOD 1 (PHOG+SVM).

Algorithm 1: Clustering based sequence quantisation algorithm

Data: Shape vector P_N from all the images of a sequence.

Normalise P_N.

Compute K-means on P_N^n for m cluster centers and i iterations.

Search one nearest neighbour P_N for each cluster m.

Sort the nearest neighbours.

[2] and static facial expression analysis [25], [32]. In [25] and [32], PHOG descriptors have been used for static facial expression analysis. At the start the canny edge detector is applied to the cropped face. Then the face is divided into spatial grids at all pyramid levels. After this a $3x3$ Sobel mask is applied to the edge contours for calculating the orientation gradients. Then the gradients of each grid are joined together at each pyramid level. There is an option for two orientation ranges, [0-180] and [0-360]. In [25], [0-360] orientation range perform better then [0-180]. In our experiments, we use number of pyramids L = 3, the bin size N = 8 and the orientation range is [0-360].

B. Appearance feature extraction using LPQ

The Local binary patterns (LBP) family of descriptors (LBP [33], LBP-TOP [23], LPQ [3] and LPQ-TOP [24]) have been extensively used for texture analysis, static and temporal facial expression analisys and face recognition. We use the LPQ (Local Phase Quantization) appearance descriptor. Though LPQ-TOP [24] has been proposed for temporal data analayis, we do not have labeling of an onset, apex and offset in the database in our experiments and hence we use LPQ only. LPQ is based on computing the short-term fourier transform (STFT) on local image window. At each pixel the local Fourier coefficients are computed for four frequency points. Then the signs of the real and imaginary part of each coefficient is quantised using a binary scalar quantiser, for calculating the phase information. The resultant eight bit binary coefficients are then represented as integers using binary coding. This step is similar to the histogram construction step in LBP. In the end, we get a 256-dimensional feature vector. In our experiments, we divided the cropped face of size 60x60 into four blocks. This gave us a vector dimension of 1024 for an image and 6144 for an image sequence where the number of cluster centres m = 6.

Emotion	PI-BL	PS-BL	PO-BL	PI-M1	PS-M1	PO-M1	PI-M2	PS-M2	PO-M2	PI-M3	PS-M3	PO-M3
Anger	0.86	0.92	0.89	0.714	0.30	0.518	0.857	0.846	0.851	**0.928**	**1.0**	**0.962**
Fear	0.07	0.40	0.20	0.666	1.0	0.80	**0.333**	**0.80**	**0.520**	0.0	**0.80**	0.32
Joy	0.70	0.73	0.71	0.60	0.454	0.548	0.70	0.545	0.645	**0.80**	**0.636**	**0.741**
Relief	0.31	0.70	0.46	0.687	1.0	0.807	0.687	**1.0**	0.807	**0.75**	**1.0**	**0.846**
Sadness	0.27	0.90	0.52	0.667	0.70	0.68	0.666	**1.0**	0.80	0.666	**1.0**	**0.80**
Average	0.44	0.73	0.56	0.667	0.69	0.67	**0.648**	0.838	0.724	0.629	**0.887**	**0.734**

TABLE II

ACCURACY COMPARISON AMONG BASELINE(PI-BL, PS-BL, PO-BL), METHOD 1(PI-M1, PS-M2, PO-M3), METHOD 2(PI-M2, PS-M2, PO-M2), METHOD 3(PI-M3, PS-M3, PO-M3). HERE, PI - PERSON INDEPENDENT PARTITION, PS - PERSON SPECIFIC PARTITION AND PO - PERSON OVERALL.

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	4	0	0	0	0
Fear	8	10	3	0	2
Joy	1	0	5	0	0
Relief	0	0	3	10	1
Sadness	0	0	0	0	7

TABLE III

CONFUSION MATRIX FOR PERSON SPECIFIC TEST FOR OUR METHOD 1 (PHOG+SVM).

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	14	3	7	0	2
Fear	9	20	3	2	2
Joy	3	2	17	0	1
Relief	1	0	4	21	3
Sadness	0	0	0	3	17

TABLE IV

OVERALL CONFUSION MATRIX FOR OUR METHOD 1 (PHOG+SVM).

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	12	7	3	2	5
Fear	2	5	0	1	0
Joy	0	2	14	0	0
Relief	0	0	3	11	0
Sadness	0	1	0	2	10

TABLE V

CONFUSION MATRIX FOR PERSON INDEPENDENT TEST FOR OUR METHOD 2 (PHOG+LPQ+SVM).

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	11	0	0	0	0
Fear	2	8	3	0	0
Joy	0	0	6	0	0
Relief	0	0	2	10	0
Sadness	0	2	0	0	10

TABLE VI

CONFUSION MATRIX FOR PERSON SPECIFIC TEST FOR OUR METHOD 2 (PHOG+LPQ+SVM).

III. EXPERIMENTS

A. GEMEP-FERA dataset

The GEMEP-FERA dataset consists of recordings of 10 actors displaying a range of expressions, while uttering a meaningless phrase, or the word Aaah. There are 7 subjects in the training data, and 6 subjects in the test set, 3 of which are not present in the training set. The training set contains 155 image sequences and the testing contains 134 image sequences. The test contains six actors, three of them are also present in the training data. This enables testing of the method on both person independent and person specific settings. There are in total five emotion categories in the dataset: *anger*, *fear*, *joy*, *relief* and *sadness*. The baseline method was provided pre computed. The method uses LBP for feature extraction. It computes LBP feature for all the frames and then classifies each frame individually. Then maximum voting mechanism is used for deciding the final emotion category of the video sequence. The average person-specific and person-independent classification accuracy of the method are 0.73 and 0.44 respectively and the overall baseline system accuracy is 0.56.

We experimented with three methods: a) PHOG features with SVM classification, b) PHOG+LPQ features with SVM classification and c) PHOG+LPQ with LMNN classification. First we track the face using CLM and then apply K-means clustering to the shape paramters. In our experiments the number of clusters centers m = 6, this was chosen empirically. Then we apply kNN and find the nearest neigbour to the computed cluster centers. Post this we apply the Viola-Jones face detector to the closest images. The cropped face's size was set to 60x60. Then for method 1, we compute the PHOG descriptors, here number of pyramids L = 3, the orientation range is [0-360]. PHOG descriptors for the images (6 images in this case) are then concatenated. In total we have 6800 dimension vector for eacch sequence. We then applied PCA to the data and 98% of the variance is maintained.

For classification we trained a non-linear SVM model [5]. For parameter selection we used ten fold cross-validation. The accuracy for method 1 is 0.66 for person independent partition, 0.69 for the person-dependent partition and 0.67 as overall accuracy. In this method we are using shape features only. Tables I, III and IV are the confusion matrix for method 1 for person independent, person dependent and overall partitions.

Next, in method 2, we added appearance features in the form of LPQ. In the literature LPQ have shown to perform better [3] than LBP and are invariant to blur and illumination up to some extent. The LPQ were calculated on the 60x60 cropped face. For LPQ, 6144 features were

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	23	7	3	2	5
Fear	4	13	3	1	0
Joy	0	2	20	0	0
Relief	0	0	5	21	0
Sadness	0	3	0	2	20

TABLE VII

OVERALL CONFUSION MATRIX FOR OUR METHOD 2 (PHOG+LPQ+SVM).

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	13	12	0	3	5
Fear	0	0	0	0	0
Joy	0	2	16	0	0
Relief	0	0	4	12	0
Sadness	1	1	0	1	10

TABLE VIII

CONFUSION MATRIX FOR PERSON INDEPENDENT TEST FOR OUR METHOD 3 (PHOG+LPQ+LMNN).

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	13	0	0	0	0
Fear	0	8	2	0	0
Joy	0	0	7	0	0
Relief	0	0	2	10	0
Sadness	0	2	0	0	10

TABLE IX

CONFUSION MATRIX FOR PERSON SPECIFIC TEST FOR OUR METHOD 3 (PHOG+LPQ+LMNN).

Emotion	Anger	Fear	Joy	Relief	Sadness
Anger	26	12	0	3	5
Fear	0	8	2	0	0
Joy	0	2	23	0	0
Relief	0	0	6	22	0
Sadness	1	3	0	1	20

TABLE X

OVERALL CONFUSION MATRIX FOR OUR METHOD 3 (PHOG+LPQ+LMNN).

generated per sequence. These are then concatenated with the PHOG features. We again apply PCA and retain 98% of the variance. Similar to method 1, in method 2 we train a SVM with RBF kernel and the parameters are selected with ten fold cross validation. Method 2 outperforms method 1 and we get a performance increase of approximately 6.4%. This is due to the addition of the appearance features. The classification performance for method 2 is as follows: for person independent it is 0.648, for the person-dependent partition it is 0.838 and the overall accuracy is 0.724. Tables V, VI and VII are the confusion matrices for method 2 for person independent, person dependent and overall partitions.

Further, we also experimented with distance learning method LMNN. Nowadays, distance learning methods have caught a lot of attention as they produce good classification results. LMNN learns a Mahanabolis distance metric over the labelled training set. With the same features as in method 2, we learn a Mahanabolis matrix in LMNN. Method 3 gives the best performance out of all the methods discussed here. For person independent partition it it 0.629, for person dependent partition it is 0.887 and the overall accuracy is 0.734. Method 3 performs better than Method 2 in the person-dependent and overall partition but less well in person independent. Also, for the emotion category *fear*, the accuracy is less than Method 2. Table VIII, describes the confusion matrix for person independent results for method 3. Here, each row is the true class and the columns are the predicted values. Similarity, Table IX describes the confusion matrix for person specific results for method 3 and Table X describes the overall all performance confusion matrix for method 3.

Table II, compares the four methods: baseline, method 1, method 2 and method 3 for their performance in person independent, person dependent and overall partitions. Figure 3, compares the overall performance of the four method for the five emotion classes.

IV. CONCLUSIONS AND FUTURE WORK

We presented a novel method for automatic emotion recognition. We used unsupervised clustering on normalised shape vectors for choosing key frames. For capturing shape information we used, PHOG features and for appearance we use the recently proposed LPQ features. We test our method for person independent and person dependent classification performance on the GEMEP-FERA data set. The data set is a challenging data base with actors posing various emotions while uttering sentences. The proposed method performs better then the baseline results. For future work we want to explore methods for feature selection such via boosting and mutual information. Also we will experiment on other databases such as the MMI database [21] and the Cohn-Kanade database [20] so as to check the generic performance of the proposed method.

REFERENCES

[1] M. Valstar, B. Jiang, M. Mehu, M. Pantic, and S. Klaus, “The first facial expression recognition and analysis challenge,” in *Automatic Face and Gesture Recognition*, 2011.

[2] A. Bosch, A. Zisserman, and X. Munoz, “Representing shape with a spatial pyramid kernel,” in *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2007.

[3] V. Ojansivu and J. Heikkil, “Blur insensitive texture classification using local phase quantization,” in *Image and Signal Processing*, ser. Lecture Notes in Computer Science, 2008.

[4] T. JBaenziger and K. R. Scherer, “Introducing the geneva multimodal emotion portrayal (gemep) corpus,” in *Blueprint for affective computing: A sourcebook*, T. B. K. R. Scherer and E. Roesch, Eds. Oxford, England: Oxford University Press.

[5] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[6] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” *JMLR*, vol. 10, pp. 207–244, 2009.

[7] J. A. Russell, J. A. Bachorowski, and J. M. Fernandez-Dols, “Facial and vocal expressions of emotion.” *Annu Rev Psychol*, vol. 54, pp. 329–349, 2003.

[8] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, "Emotion Analysis in Man-Machine Interaction Systems," in *Procceedings of the First International Workshop on Machine Learning for Multimodal Interaction*. Springer, 2005, pp. 318–328.

[9] J. Bassili, "Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face," *J Personality Socical Psychol*, 1979.

[10] Z. Ambadar, J. Schooler, and J. Cohn, "Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions," *Psychological Science*, 2005.

[11] Y. Yacoob and L. Davis, "Computing spatio-temporal representations of human faces," in *In CVPR*. IEEE Computer Society, 1994, pp. 70–75.

[12] M. J. Black, D. J. Fleet, and Y. Yacoob, "A framework for modeling appearance change in image sequences," in *ICCV*, 1998, pp. 660–667.

[13] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE TPAMI*, vol. 21, no. 10, pp. 974–989, 1999.

[14] P. Ekman and W. Friesen, "The facial action coding system: A technique for the measurement of facial movement," in *Consulting Psychologists*, 1978.

[15] S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. de la Torre, and J. Cohn, "AAM derived face representations for robust facial action recognition," in *IEEE AFGR*, 2006.

[16] J. Whitehill, G. Littlewort, I. R. Fasel, M. S. Bartlett, and J. R. Movellan, "Toward practical smile detection," *IEEE TPAMI*, vol. 31, no. 11, pp. 2106–2111, 2009.

[17] M. Pantic, I. Patras, and L. Rothkruntz, "Facial action recognition in face profile image sequences," in *Proceedings of the 2002 IEEE International Conference on Multimedia and Expo, 2002. ICME '02.*, vol. 1, 2002, pp. 37 – 40 vol.1.

[18] M. Pantic, I. Patras, and M. Valstar, "Learning spatiotemporal models of facial expressions," in *Int'l Conf. Measuring Behaviour 2005*, August 2005, pp. 7–10. [Online]. Available: http://pubs.doc.ic.ac.uk/Pantic-MB05/

[19] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE TPAMI*, vol. 23, no. 3, pp. 257 –267, Mar. 2001.

[20] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. Int. Conf. on Automatic Face and Gesture Recognition (FG'00)*. Grenoble, France: IEEE, 2000, pp. 46–53.

[21] M. Stewart Bartlett, G. Littlewort, M. Frank, C. Lainscsek1, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior," in *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 223–230.

[22] M. Valstar and M. Pantic, "Fully automatic facial action unit detection and temporal analysis," in *IEEE Int'l Conf. on Computer Vision and Pattern Recognition 2006*, vol. 3, May 2006. [Online]. Available: http://pubs.doc.ic.ac.uk/Pantic-CVPR06-2/

[23] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE TPAMI*, vol. 29, no. 6, pp. 915–928, 2007.

[24] B. Jiang, M. Valstar, and M. Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in *IEEE AFGR*, 2011.

[25] Z. Li, J. Imai, and M. Kaneko, "Facial-component-based bag of words and phog descriptor for facial expression recognition," in *Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, ser. SMC'09, 2009.

[26] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *In Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.

[27] J. M. Saragih, S. Lucey, and J. Cohn, "Face alignment through subspace constrained mean-shifts," in *International Conference of Computer Vision (ICCV)*, September 2009.

[28] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR (1)*, 2001.

[29] P. Yang, Q. Liu, and D. Metaxas, "Similarity features for facial event analysis," in *Proceedings of the 10th European Conference on Computer Vision: Part I*, ser. ECCV '08, 2008.

[30] F. Wallhoff, "Facial expressions and emotion database," 2006, `http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html`.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition*, vol. 2, June 2005, pp. 886–893.

[32] Y. Bai, L. Guo, L. Jin, and Q. Huang, "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition," in *Proceedings of the 16th IEEE international conference on Image processing*, ser. ICIP'09, 2009.

[33] T. Ojala, M. Pietikinen, and T. Menp, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE TPAMI*, vol. 24, no. 7, pp. 971–987, 2002.